

Averages Without Extremes In Excel

October 28, 2003

By Neil J. Rubenking

I have a column in Excel that contains account balances. I would like to find the average balance after throwing out the one or two highest and one or two lowest values. I have not found a clean way to keep track of the count while averaging only the values that meet my criteria. Is there a way to choose how many values to trim from the top and bottom before calculating the average?

Richard Wolff

Editor's Note: If you want to omit the same number of items from the top and bottom, or omit a certain percentage from top and bottom, you can use the TRIMMEAN() function. TRIMMEAN() takes as its arguments a range of values and a percentage between 0 and 1. If you specify 0.1, ten percent of the values will be omitted, taken equally from the top and bottom. In the example, you could omit two values, one each from the top and bottom, using the simple formula TRIMMEAN(Bal, 2/COUNT(Bal)). To omit two each from top and bottom, you'd use TRIMMEAN(Bal, 4/COUNT(Bal)).

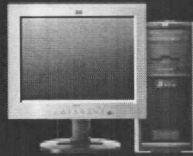
Suppose your balances are in a range named Bal that starts in cell A1 and extends down column A. If you want to discard only the highest and lowest values, you could do that with a formula like this: =(SUM(Bal)-SMALL(Bal,1)-LARGE(Bal,1))/(COUNT(Bal)-2). But we've found no easy way to go from this to a formula that lets you discard a variable number of items from the high and low ends. Instead, we worked up a second range of cells that indicate whether the adjacent cells should be included in the average.

Suppose the number of low values to be dropped is in cell D1, the number of high values to be dropped is in cell D2, and column B is empty. In cell B1, enter this formula: =AND(A1>SMALL(Bal,D\$1),A1

Unfortunately, that system breaks down when the value in D1 or D2 is 0 (that is, if you cut off no items at the high or low end). If your spreadsheet must handle that eventuality, change cell B1 to this lengthy formula: =AND(IF(D\$1>=0,A1>SMALL(Bal,D\$1),TRUE),IF(D\$2>=0,A1<LARGE(Bal,D\$2),TRUE)). Then copy the formula throughout the remainder of the Temp range. You'll still get a #DIV/0 error if you cut off so many that none remain, but that's not unreasonable.

Copyright (c) 2003 Ziff Davis Media Inc. All Rights Reserved.

Arial 10 B I U \$ % , +.0 -.00								
Reply with Changes... End Review...								
C16 $\text{=SUMIF(Temp,TRUE,Bal)/COUNTIF(Temp,TRUE)}$								
	A	B	C	D	E	F	G	H
1	0	FALSE	Cut low	10				
2	1	FALSE	Cut high	3				
3	10	FALSE						
4	20	FALSE						
5	30	FALSE						
6	40	FALSE						
7	50	FALSE						
8	60	FALSE						
9	70	FALSE						
10	80	FALSE						
11	90	TRUE						
12	100	FALSE						
13	10000	FALSE						
14	20000	FALSE						
15								
16			90					
17								



Get more power.
Get HP workstations with dual Intel® Xeon™ processors
» I want to get more

Better Averages Without Extremes

January 20, 2004

By Neil J. Rubenking

In your issue of October 28 ("Averages Without Extremes in Excel"), you presented a technique for calculating the average of a range of cells while omitting one or more of the highest or lowest values. Your technique required the addition of an extra column. I came up with an array formula that doesn't require any extra cells. Let's assume the same situation as the original, with the data cells in a range named Bal, the number of low items to cut off in D1, and the number of high items to cut off in D2. Enter this formula and press Ctrl-Shift-Enter.

```
=AVERAGE(SMALL(Bal,ROW(INDIRECT (D1+1&"": "&COUNT(Bal)-D2))))
```

Let's look at the formula from the inside out. Suppose there are 20 items in the range named Bal, that D1 (the low cutoff) is 3, and that D2 (the high cutoff) is 2. The expression D1+1&"": "&COUNT(Bal)-D2 resolves to 4&"": "&18, which is equivalent to 4:18. The INDIRECT function converts this text into a reference that can in turn be passed to the ROW function. ROW normally returns the row number of a cell passed to it or the row number of the first cell in a range. But because we have created an array formula, ROW returns an array representing all row numbers in that range: {4;5;6;7;8;9;10;11;12;13;14;15;16;17;18}.

The array in turn is passed as the second argument to the SMALL function, which would then become SMALL(Bal, {4;5;6;7;8;9;10;11;12;13;14;15;16;17;18}). Again, because this is an array formula, SMALL returns an array containing the fourth-smallest value, the fifth-smallest value, and so on to the eighteenth smallest value. The array contains all values in the range named Bal except the lowest three and the highest two, precisely the values we were looking for. The AVERAGE function returns the average of the values in this array.

Shane Devenshire

Many readers wrote in with ideas on better ways to calculate an average. Some erroneously assumed that the data in the Bal range was sorted. Some assumed (as we did in the October 28 article) that the Bal range contained only unique values. If more than one cell contained the Nth smallest or largest value, such formulas would return the wrong result. A few readers pointed out the TRIMMEAN function, which can trim a specified percentage of values equally from top and bottom. But the array formula shown here is by far the most elegant of those solutions.

You do need to ensure that the low cutoff and the high cutoff don't overlap, or else you'll get a wildly incorrect result. Suppose the low cutoff is 17 and the high cutoff is 16, which is clearly impossible, since there are only 20 items. The range passed to the ROW function will be 18:4, which is treated exactly the same as 4:18. The result will be precisely the same as if you had chosen to cut three low values and two high values.

For more on array formulas, see "Excel's Array Formulas".

Copyright (c) 2004 Ziff Davis Media Inc. All Rights Reserved.